Review of Economic Studies (1980) XLVII, 275–291
0034-6527/80/00130275$02.00


# Testing Linear *versus* Logarithmic Regression Models

GWYN ANEURYN-EVANS
*University of Cambridge*

and

ANGUS DEATON
*University of Bristol*

## INTRODUCTION

One of the problems most frequently encountered by the applied econometrician is the choice between logarithmic and linear regression models. Economic theory is rarely of great help although there are cases where one or other specification is clearly inappropriate; for example, in demand analysis constant elasticity specifications are inconsistent with the budget constraint. Nor are standard statistical tests very useful; $R^2$ statistics are not commensurable between models with dependent variables in levels and in logarithms and the comparison of likelihoods has no firm basis in statistical inference. In this paper, we develop a practical text based upon Cox's ((1961) and (1962)) procedure for testing separate families of hypotheses; the work is thus an extension of earlier econometric applications of Cox's test to single equation linear regressions in Pesaran (1974) and to many equation non-linear regression in Pesaran and Deaton (1978). The test we develop here is applicable to two competing single-equation models, one of which explains the *level* of a variable up to an additive error, the other of which explains its *logarithm*, again up to an additive error. Hence, in terms of the levels of the variables, we are testing for *multiplicative* versus *additive* errors, and it is this which differentiates this paper from the earlier work in which an additive error was always assumed. We shall also allow, as in the earlier papers, the deterministic parts of the regressions to be linear or non-linear and to have the same or different independent variables; it is thus possible to test for functional form and specification in a very general way.

Section 1 of the paper defines the problem and derives the test statistics. The formulae allow the calculation of two statistics, $N_0$ and $N_1$ say, the first of which is asymptotically distributed as $N(0, 1)$ if the logarithmic specification is correct, the second, for all practical purposes, as $N(0, 1)$ if the linear model is true. Section 2 discusses problems associated with the calculation of the statistics and shows how they can be surmounted. Section 3 presents the results of Monte-Carlo experiments designed to evaluate the potential of the test in practice. We investigate, in particular, the shape of the actual distributions of $N_0$ and $N_1$ in samples of sizes 20, 40 and 80 as well as comparing the performance of the Cox procedure with that of the likelihood ratio test, as proposed by Sargan (1964). Finally, we offer some evidence of the ability of the procedure to detect total misspecification when *neither* of the hypotheses is true. Section 4 contains a summary and conclusions.

The general issues of statistical inference raised by the use of the Cox procedure in econometrics as well as alternative testing procedures have already been widely discussed, see Pesaran and Deaton (1978), Quandt (1974) and Amemiya (1976). In this case, however, there exists one very obvious alternative procedure. This is to specify the model, not in levels or in logarithms, but via the Box-Cox transform; hence, the dependent variable is $(y^{\alpha}-1)/\alpha$, so that with $\alpha=1$, the regression is linear, with $\alpha=0$, it is logarithmic, these cases being only two possibilities out of an infinite range as $\alpha$ varies. The general model can be estimated by grid search or by non-linear maximization of the likelihood and a maximum likelihood estimate for $\alpha$ obtained. The values of 0 and 1 can then be compared using a conventional likelihood ratio test. A number of comments may be made. Note that the problem is now rather different from the one originally posed. We no longer have a discrete choice between $\alpha=0$ and $\alpha=1$ but instead have to choose between one or both of these and the maximum-likelihood estimate, $\hat{\alpha}$, say. One possible difficulty is that the investigator may only be interested in linear or logarithmic forms so that a value for $\hat{\alpha}$ of 0·73, for example, may not be very useful. It is our impression that many econometricians would simply adopt a rule which chooses the linear form when $\hat{\alpha}>0{\cdot}5$ and the log-linear $\hat{\alpha}<0{\cdot}5$. If so, a straightforward comparison of likelihoods between the two cases is much more appropriate, and, in Section 3, we shall examine the performance of such a rule. However, it may well be that the investigator has no prior preference for either the logarithmic or linear forms so that any value for $\hat{\alpha}$ is perfectly acceptable. In this case, the correct procedure is to estimate the Box-Cox model and use the value of $\hat{\alpha}$ which emerges; no problem of choice or of inference arises. The final possibility is that the true model may not be included even in the general form. In this case, estimation of $\alpha$ and comparison with 0 and 1 will be of no help; the whole experiment is being conducted within a false maintained hypothesis and is meaningless. The Cox procedure, however, encompasses the possibility of rejecting both models which is, in this case, the only correct decision. Whether in fact it will do so is a matter for empirical investigation and we shall return to the question in Section 3.

## 1. THE DERIVATION OF THE TEST STATISTICS

### **1.1.** *Specification of the Models*

There are two models to be compared, $H_0$ and $H_1$. These are defined by

$$H_0\colon \ln y_t = f_t(x,\theta_0)+\varepsilon_{0t}, \qquad \ldots(1)$$

$$H_1\colon y_t = g_t(z,\theta_1)+\varepsilon_{1t}, \qquad \ldots(2)$$

where $y_t$ is the $t$th observation on the dependent variable, $t=1,\ldots,T$; $\theta_0$ and $\theta_1$ are $k_0$ and $k_1$ vectors of parameters, and $x$ and $z$ are vectors of independent variables. There is no restriction on the variables which appear in $x$ and $z$; they may be the same, different or transformations of one another. The functions $f_t$ and $g_t$ may or may not be linear. For $\varepsilon_{0t}$, we assume

$$\varepsilon_{0t}\sim N(0,\sigma_0^2). \qquad \ldots(3)$$

It is not, however possible to assume $\varepsilon_{1t}\sim N(0,\sigma_1^2)$ since, if this were so, there would always be a finite probability of $y_t$ becoming non-positive. If so, $H_0$ must be false since the logarithm does not exist, and the problem of inference is a trivial one. Consequently, if it is possible to give serious consideration to $H_0$, the distribution of $\varepsilon_{1t}$ must be such as to ensure that $y$ is positive. Various distributions could be used which would meet this criterion. For example, we could follow Amemiya (1973) and truncate $\varepsilon_{1t}$ so that $y_t$ is given by (2) if the right hand side is greater than some positive number, $a$, say, and equal to $a$ otherwise. Alternatively and in some respects more simply, we can truncate the distribution of $\varepsilon_{1t}$ at a fixed number of standard deviations from zero. Here we use the normal distribution and it is convenient to truncate symmetrically, i.e. if $f_N(x;\sigma^2)$ is the p.d.f. of an $N(0,\sigma^2)$ variable and if $\pi(\varepsilon_{1t})$ is the p.d.f. of $\varepsilon_{1t}$, we write

$$\begin{aligned}\pi(\varepsilon_{1t}) &= \alpha(k) f_N(\varepsilon_{1t};\sigma_1^2) \qquad && |\varepsilon_{1t}|\leqq k\sigma_1\\ &=0 && |\varepsilon_{1t}|>k\sigma_1 \qquad \ldots(4a)\end{aligned}$$

where

$$\alpha(k) = \left\{ \int_{-k}^{k} (2\pi)^{-\frac{1}{2}} e^{-t^2/2} dt \right\}^{-1} \qquad \ldots(4b)$$

The fact that $H_0$ is to be seriously considered will be taken to imply that $k$ can be set *a priori* at some large value ($k \geqq 6$ say). For values of $k$ of the size indicated $\alpha$ is so close to unity that, for many purposes, we can assume $\varepsilon_{1t} \sim N(0, \sigma_1^2)$. If, on the other hand, we are not prepared to assert that $g_t(z, \theta_1)$ is always at least 6 to 8 equation standard errors above zero, then it is not sensible to regard both $H_0$ and $H_1$ as possible.

Let us write $\alpha_0$ and $\alpha_1$ for the extended parameter vectors of the two models, so that $\alpha_0' = (\theta_0', \sigma_0^2)$ and $\alpha_1' = (\theta_1', \sigma_1^2)$. Denote the log likelihood functions of $H_0$ and $H_1$ by $L_0(\alpha_0)$ and $L_1(\alpha_1)$ respectively and by $\hat{L}_{10}$ the log of the maximum likelihood ratio. Then, if $H_0$ is true, we must calculate

$$T_0 = \hat{L}_{10} - T\left\{ \underset{T\to\infty}{\text{plim}_0} \left( \frac{\hat{L}_{10}}{T} \right) \right\}_{\alpha_0 = \hat{\alpha}_0} \qquad \ldots(5)$$

where $\text{plim}_0$ denotes the probability limit when $H_0$ is true and $\hat{}$ denotes a maximum likelihood estimate. If $L_{10} = L_0(\alpha_0) - L_1(\alpha_{10})$, where $\alpha_{10} = \text{plim}_0\, \hat{\alpha}_1$, then Cox (1962) has shown that, if $H_0$ is true, $T_0$ is asymptotically normally distributed with mean zero and variance $V_0(T_0)$, given by

$$V_0(T_0) = V_0(L_{10}) - \frac{1}{T} \eta' Q^{-1} \eta, \qquad \ldots(6)$$

where $Q$ is the asymptotic information matrix of $H_0$,

$$Q = -\text{plim}_0 \frac{1}{T} \frac{\partial^2 L_0}{\partial\alpha_0 \partial\alpha_0'} \qquad \ldots(7)$$

and

$$\eta = T \frac{\partial\, [\text{plim}_0\, (\hat{L}_{10}/T)]}{\partial\alpha_0}. \qquad \ldots(8)$$

If $H_1$ is true, similar expressions yield $T_1$ and $V_1(T_1)$.

The two log likelihood functions are given by

$$L_0(\alpha_0) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma_0^2 - \frac{\sum \varepsilon_{0t}^2}{2\sigma_0^2} - \sum \ln y_t \qquad \ldots(9)$$

and

$$L_1(\alpha_1) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma_1^2 - \frac{\sum \varepsilon_{1t}^2}{2\sigma_1^2} + T \ln \alpha(k) \qquad \ldots(10)$$

where the extra sum in (9) is the log of the product of the Jacobians. Maximum likelihood estimates $\hat{\theta}_0$, $\hat{\sigma}_0^2$, $\hat{\theta}_1^2$, $\hat{\sigma}_1^2$ satisfy

$$\hat{\sigma}_0^2 = \frac{1}{T} \sum_t \{\ln y_t - f_t(\hat{\theta}_0)\}^2, \qquad \ldots(11)$$

$$\sum_t \frac{\partial f_t(\hat{\theta}_0)}{\partial \theta_{0j}} \{\ln y_t - f_t(\hat{\theta}_0)\} = 0, \qquad j = 1, \ldots, k_0 \qquad \ldots(12)$$

$$\hat{\sigma}_1^2 = \frac{1}{T} \sum_t \{y_t - g_t(\hat{\theta}_1)\}^2, \qquad \ldots(13)$$

$$\sum_t \frac{\partial g_t(\hat{\theta}_1)}{\partial \theta_{1j}}\{y_t - g_t(\hat{\theta}_1)\} = 0, \qquad j = 1, \ldots, k_1 \qquad \ldots(14)$$

where, for convenience, the $x$ and $z$ arguments of $f$ and $g$ have been suppressed.

The maximized log likelihood ratio $\hat{L}_{10}$ is thus

$$\hat{L}_{10} = \frac{T}{2}\ln\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2}\right) - \sum_t \ln y_t - T\ln\alpha(k). \qquad \ldots(15)$$

Note that $\hat{L}_{10}$ is itself a possible candidate for discriminating between $H_0$ and $H_1$ as has been suggested by Sargan (1964). The models are, of course, not nested, so that we do not have the usual justification for the statistic. Nevertheless, it is simple to calculate and, to the extent that likelihoods can be regarded as fundamental measures of plausibility, see in particular Edwards (1972), $\hat{L}_{10}$ is a natural quantity to inspect.

### 1.2. *Calculation of* $T_0$

Having laid this groundwork, we begin with the easier case, when $H_0$ is true. Writing $\alpha_{10}$ for $\text{plim}_0\, \hat{\alpha}_1$ so that $\sigma_{10}^2$ is $\text{plim}_0\, \hat{\sigma}_1^2$, from (15) and (1)

$$\text{plim}_0\left(\frac{\hat{L}_{10}}{T}\right) = \tfrac{1}{2}\ln\frac{\sigma_{10}^2}{\sigma_0^2} - \text{plim}_0\left\{\frac{1}{T}\sum f_t(\theta_0)\right\} - \ln\alpha(k). \qquad \ldots(16)$$

Hence, writing $\hat{\sigma}_{10}^2$ for the ML estimate of $\sigma_{10}^2$, $T_0$ is given by

$$T_0 = \frac{T}{2}\ln\frac{\hat{\sigma}_1^2}{\hat{\sigma}_{10}^2} - \sum_t\{\ln y_t - f_t(\hat{\theta}_0)\}. \qquad \ldots(17)$$

Note that in many applications, for example when $f_t$ is linear, the sum on the right hand side of (17) will be zero. The maximum likelihood estimate $\hat{\sigma}_{10}$, and hence $\hat{\sigma}_{10}^2$, can be calculated by solving

$$E_0\left(\frac{\partial L_1(\alpha_{10})}{\partial \alpha_{10}}\right) = 0. \qquad \ldots(18)$$

But

$$L_1(\alpha_{10}) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma_{10}^2 - \frac{1}{\sigma_{10}^2}\sum_t\{y_t - g_t(\theta_{10})\}^2 + T\ln\alpha(k) \qquad \ldots(19)$$

Hence

$$E_0\left(\frac{\partial L_1(\alpha_{10})}{\partial \sigma_{10}^2}\right) = -\frac{T}{2\sigma_{10}^2} + \frac{1}{2\sigma_{10}^4}E_0\sum_t\{y_t - g_t(\theta_{10})\}^2. \qquad \ldots(20)$$

This can be further evaluated by writing $y_t = e^{f_t(\theta_0)+\varepsilon_{0t}}$. In this and the sequel, the following expectations are useful

$$E_0(y_t^r) = e^{rf_t(\theta_0)}e^{r^2\sigma_0^2/2}, \qquad \ldots(21a)$$

$$E_0(\varepsilon_{0t}y_t^r) = r\sigma_0^2 e^{rf_t(\theta_0)}e^{r^2\sigma_0^2/2}, \qquad \ldots(21b)$$

$$E_0(\varepsilon_{0t}^2 y_t^r) = r^2\sigma_0^4 e^{rf_t(\theta_0)}e^{r^2\sigma_0^2/2} + \sigma_0^2 e^{rf_t(\theta_0)}e^{r^2\sigma_0^2/2}. \qquad \ldots(21c)$$

Hence

$$\begin{aligned} &E_0[\textstyle\sum_t\{y_t^2 - 2g_t(\theta_{10})y_t + g_t(\theta_{10})^2\}] \\ &= \textstyle\sum_t\{y_{0t}^2 e^{\sigma_0^2} - 2g_t(\theta_{10})y_{0t} + g_t(\theta_{10})^2\} \\ &= \textstyle\sum_t\{y_{0t} - g_t(\theta_{10})\}^2 + \sum_t y_{0t}^2(e^{\sigma_0^2} - 1), \end{aligned} \qquad \ldots(22)$$

where $y_{0t} = e^{f_t(\theta_0)+\sigma_0^2/2} = E_0(y_t)$, i.e. the expectation of $y_t$ if the logarithmic model is correct. Hence, combining (18), (20) and (22), $\hat{\sigma}_{10}^2$ is given by

$$\hat{\sigma}_{10}^2 = \frac{1}{T}\sum_t \{\hat{y}_{0t} - g_t(\hat{\theta}_{10})\}^2 + \frac{1}{T}\sum_t \hat{y}_{0t}^2(e^{\hat{\sigma}_0^2} - 1), \qquad \ldots(23)$$

where $\hat{y}_{0t} = e^{f_t(\hat{\theta}_0)+\hat{\sigma}_0^2/2}$. This requires the estimate $\hat{\theta}_{10}$ which is given by consideration of

$$E_0\left(\frac{\partial L_1(\alpha_{10})}{\partial \theta_{10}}\right) = 0.$$

This is straightforward; $\hat{\theta}_{10j}$, $j = 1, \ldots, k_1$ must satisfy

$$\sum_t \frac{\partial g_t(\hat{\theta}_{10})}{\partial \theta_{10j}}\{\hat{y}_{0t} - g_t(\hat{\theta}_{10})\} = 0. \qquad \ldots(24)$$

Clearly, $\hat{\theta}_{10}$ is defined by regressing the expected value of $y$ from $H_0$, i.e. $y_0$, on $g(z, \theta_1)$. Equations (12), (13), (17), (23) and (24) completely define $T_0$.

### 1.3. *Calculation of* $V_0(T_0)$

The information matrix $Q$ may be straightforwardly calculated by differentiation. Writing $M_f$ for the matrix whose $(i, j)$th term is given by

$$(M_f)_{ij} = \operatorname*{plim_0}_{T\to\infty} \frac{1}{T}\sum_t \frac{\partial f_t(\theta_0)}{\partial \theta_{0i}}\frac{\partial f_t(\theta_0)}{\partial \theta_{0j}}, \qquad \ldots(25)$$

we have

$$Q^{-1} = \begin{pmatrix} \sigma_0^2 M_f^{-1} & 0 \\ 0 & 2\sigma_0^4 \end{pmatrix}. \qquad \ldots(26)$$

To derive the elements of $\eta$, we differentiate (16) w.r.t. $\theta_{0j}$ and $\sigma_0^2$ in turn. Thus, for $j = 1, \ldots, k_0$

$$\frac{\partial \operatorname{plim}_0 (\hat{L}_{10/T})}{\partial \theta_{0j}} = \frac{1}{2\sigma_{10}^2}\frac{\partial \sigma_{10}^2}{\partial \theta_{0j}} - \operatorname{plim}_0 \frac{1}{T}\sum_t \frac{\partial f_t(\theta_0)}{\partial \theta_{0j}}$$

But, from (23), and since

$$\partial y_{0t}/\partial \theta_{0j} = y_{0t}\frac{\partial f_t(\theta_0)}{\partial \theta_{0j}}$$

$$\frac{\partial \sigma_{10}^2}{\partial \theta_{0j}} = \frac{2}{T}\sum_t \{y_{0t} - g_t(\theta_{10})\}\left\{y_{0t}\frac{\partial f_t(\theta_0)}{\partial \theta_{0j}} - \sum_k \frac{\partial g_t(\theta_{10})}{\partial \theta_{0k}}\cdot\frac{\partial \theta_{10k}}{\partial \theta_{10j}}\right\} + \frac{2}{T}(e^{\sigma_0^2} - 1)\sum_t y_{0t}^2 \frac{\partial f_t(\theta_0)}{\partial \theta_{0j}}.$$

But, from (24) we can eliminate the terms in $\partial g_t(\theta_{10})/\partial \theta_{10k}$, so that, combining, for $j = 1, \ldots, k_0$

$$\eta_j = \frac{1}{\sigma_{10}^2}\sum_t y_{0t}\frac{\partial f_t(\theta_0)}{\partial \theta_{0j}}\{e^{\sigma_0^2}y_{0t} - g_t(\theta_{10})\} - \sum_t \frac{\partial f_t(\theta_0)}{\partial \theta_{0j}}. \qquad \ldots(27)$$

Similarly, differentiating (16) w.r.t. $\sigma_0^2$ and using (8),

$$\eta_{k_0+1} = \frac{1}{2\sigma_{10}^2}\sum_t y_{0t}\{2e^{\sigma_0^2}y_{0t} - g_t(\theta_{10})\} - \frac{T}{2\sigma_0^2}. \qquad \ldots(28)$$

From (9) and (19) $L_{10} = L_0(\alpha_0) - L_1(\alpha_{10})$ is given by

$$L_{10} = -\frac{T}{2}\sigma_0^2 - \frac{1}{2\sigma_0^2}\sum_t \varepsilon_{0t}^2 - \sum \ln y_t + \frac{T}{2}\sigma_{10}^2 + \frac{1}{2\sigma_{10}^2}\sum\{y_t - g_t(\theta_{10})\}^2 - T\ln\alpha(k).$$

The variance of this expression can be evaluated straightforwardly (but tediously) term by term using the expectations (21). Hence

$$V_0(L_{10})=\frac{T}{2}+T\sigma_0^2+\frac{1}{4\sigma_{10}^4}[e^{2\sigma_0^2}(e^{4\sigma_0^2}-1)\sum y_{0t}^4-4e^{\sigma_0^2}(e^{2\sigma_0^2}-1)\sum y_{0t}^3g_t(\theta_{10})$$
$$+4(e^{\sigma_0^2}-1)\sum y_{0t}^2g_t(\theta_{10}^2)]-\frac{\sigma_0^2}{\sigma_{10}^2}\sum y_{0t}[4e^{\sigma_0^2}y_{0t}-3g_t(\theta_{10})].$$

Hence, from (6), (26), (27) and (28),

$$V_0(T_0)=T\sigma_0^2+\frac{1}{4\sigma_{10}^4}[e^{2\sigma_0^2}(e^{4\sigma_0^2}-1)\sum y_{0t}^4$$
$$-4e^{\sigma_0^2}(e^{2\sigma_0^2}-1)\sum y_{0t}^3g_t(\theta_{10})+4(e^{\sigma_0^2}-1)\sum y_{0t}^2g_t(\theta_{10})^2$$
$$-\frac{\sigma_0^2}{T}\zeta' M_f^{-1}\zeta-\frac{2\sigma_0^2}{\sigma_{10}^2}\sum y_{0t}[y_{0t}e^{\sigma_0^2}-g_t(\theta_{10})]$$
$$-\frac{\sigma_0^4}{2T\sigma_{10}^4}[\sum y_{0t}(2e^{\sigma_0^2}-g_t(\theta_{10}))]^2, \qquad \ldots(29)$$

where $\zeta'=(\zeta_1,\ldots,\zeta_{k_0})$.
In practice, (29) is evaluated by replacing each of the unknown quantities $\theta_0$, $\theta_{10}$, $\sigma_0^2$, $\sigma_{10}^2$ and $M_f$ by their maximum likelihood estimates $\hat{\theta}_0$, $\hat{\theta}_{10}$, $\hat{\sigma}_0^2$, $\hat{\sigma}_{10}^2$ and $T^{-1}\sum_t(\partial f_t/\partial\theta_0)\cdot(\partial f_t/\partial\theta_0')$.

### 1.4. *Calculation of $T_1$*

In the earlier papers, Pesaran (1974) and Pesaran and Deaton (1978), there was no need to derive $T_1$ and $V_1(T_1)$ separately since the derivations for $T_0$ and $V_0(T_0)$ could be repeated with merely a change of suffix. In the present case however, the asymmetry between the models requires quite separate derivation for the case where $H_1$ is supposed to be true. In what follows, we shall frequently take expectations of expressions involving $\varepsilon_{1t}$. These cannot be evaluated analytically and we shall return in Section 2 to how they are best computed in practice. Since $\hat{L}_{01}=-\hat{L}_{10}$,

$$\text{plim}_1\left(\frac{\hat{L}_{01}}{T}\right)=-\tfrac{1}{2}\ln\frac{\sigma_1^2}{\sigma_{01}^2}+\text{plim}_1\frac{1}{T}\sum\ln y_t-\ln\alpha(k), \qquad \ldots(30)$$

where $\sigma_{01}^2=\text{plim}_1\,\hat{\sigma}_0^2$. Hence,

$$T_1=\frac{T}{2}\ln\frac{\hat{\sigma}_0^2}{\hat{\sigma}_{01}^2}+\sum_t\{\ln y_t-E_1(\ln y_t)\}, \qquad \ldots(31)$$

where the expectation $E_1(\;)$ is evaluated using the ML estimates $\hat{\theta}_1^2$ and $\hat{\sigma}_1^2$. As before, $\hat{\sigma}_{01}^2$ (and $\hat{\theta}_{01}$) are derived by solving $E_1\{\partial L_0(\alpha_{01})/\partial\alpha_{01}\}=0$. Substituting $\alpha_{01}$ in (9) and differentiating

$$\frac{\partial L_0(\alpha_{01})}{\partial\sigma_{01}^2}=-\frac{T}{2\sigma_{01}^2}+\frac{1}{2\sigma_{01}^4}\sum_t\{\ln y_t-f_t(\theta_{01})\}^2.$$

Hence, taking expectations and setting to zero,

$$\hat{\sigma}_{01}^2=\frac{1}{T}\sum_t\text{var}_1(\ln y_t)+\frac{1}{T}\sum_t\{E_1(\ln y_t)-f_t(\hat{\theta}_{01})\}^2. \qquad \ldots(32)$$

$\hat{\theta}_{01}$ is estimated similarly and must be calculated from the implicit solution of

$$\sum_t \{E_1(\ln y_t) - f_t(\hat{\theta}_{01})\} \frac{\partial f_t(\hat{\theta}_{01})}{\partial \theta_{01j}} = 0, \qquad j = 1, \ldots, k_0. \qquad \ldots(33)$$

In practice, this is simply a repetition of the estimation of $\theta_0$ using $E_1(\ln y_t)$ instead of $\ln y_t$ as dependent variable. Equations (31)–(33) are sufficient, together with (11) and (12), for the calculation of $T_1$.

**1.5.** *Calculation of* $V_1(T_1)$

We shall need the result, derived from (4) by integration by parts,

$$E(\varepsilon_{1t}^r \sigma_1^{-r}) = -\{1 - (-1)^{r-1}\} k^{r-1} \omega_k + (r-1) E(\varepsilon_{1t}^{r-2} \sigma_1^{-(r-2)})$$

where $\omega_k = (2\pi)^{-\frac{1}{2}} \alpha(k) e^{-(k^2/2)}$. Hence, the four moments of $\varepsilon_{1t}$ are given by

$$E(\varepsilon_{1t}) = 0 \qquad E(\varepsilon_{1t}^2) = \sigma_1^2 (1 - a_k)$$
$$E(\varepsilon_{1t}^3) = 0 \qquad E(\varepsilon_{1t}^4) = 3\sigma_1^4 (1 - b_k)$$

where $a_k = 2k\omega_k$, and $b_k = 2k(1 + k^2/3)\omega_k$. Note that, for the relevant range of $k$, both $a_k$ and $b_k$ are small (see the table below).

| $k$ | $a_k$ | $b_k$ |
|---|---|---|
| 6 | $0{\cdot}7 \cdot 10^{-7}$ | $0{\cdot}9 \cdot 10^{-6}$ |
| 7 | $1{\cdot}3 \cdot 10^{-10}$ | $2{\cdot}2 \cdot 10^{-9}$ |
| 8 | $0{\cdot}8 \cdot 10^{-13}$ | $1{\cdot}8 \cdot 10^{-12}$ |

Now, from (10) and the above expectations

$$E\left(\frac{1}{T} \frac{\partial \ln L_1}{\partial \theta_{1i}}\right) = 0 \qquad E\left(\frac{1}{T} \frac{\partial \ln L_1}{\partial \sigma_1^2}\right) = -\frac{T}{2\sigma_1^2} a_k$$

$$E\left(\frac{1}{T} \frac{\partial^2 \ln L}{\partial \theta_{1i} \partial \theta_{1j}}\right) = -\frac{1}{T\sigma_1^2} \sum \frac{\partial g_t}{\partial \theta_{1i}} \frac{\partial g_t}{\partial \theta_{1j}} E\left(\frac{1}{T} \frac{\partial^2 \ln L}{\partial \theta_{1i} \partial \sigma_1^2}\right) = 0$$

$$E\left(\frac{1}{T} \frac{\partial^2 \ln L_1}{\partial \sigma_1^2 \partial \sigma_1^2}\right) = -\frac{T}{2\sigma_1^4}(1 - 2a_k).$$

Hence, the information matrix $Q$ is given by

$$Q^{-1} = \begin{pmatrix} \sigma_1^2 M_g^{-1} & 0 \\ 0 & \dfrac{2\sigma_1^4}{1 - 2a_k} \end{pmatrix} \qquad \ldots(34)$$

where $M_g$ is defined analagously to $M_f$ (see (25) above).

One further point needs to be made. For the asymptotic convergence of the Cox statistic to $N(0, 1)$ it is sufficient that

$$E\left(\frac{\partial L_1}{\partial \alpha_1}\right) = 0 \qquad \ldots(35a)$$

and

$$E\left(-\frac{\partial^2 L_1}{\partial \alpha_1 \partial \alpha_1'}\right) = E\left(\left(\frac{\partial L_1}{\partial \alpha_1}\right)\left(\frac{\partial L_1}{\partial \alpha_1}\right)'\right). \qquad \ldots(35b)$$

For all practical purposes, (35a) is satisfied for the values of $k$ considered; for (35b) we find that

$$E\left(-\frac{\partial^2 L_1}{\partial\theta_{1i}\partial\theta_{1j}}\right)=E\left(\frac{\partial L_1}{\partial\theta_{1i}}\frac{\partial L_1}{\partial\theta_{1j}}\right),$$

$$E\left(-\frac{\partial^2 L_1}{\partial\theta_{1i}\partial\sigma_1^2}\right)=E\left(\frac{\partial L_1}{\partial\theta_{1i}}\frac{\partial L_1}{\partial\sigma_1^2}\right),$$

but

$$E\left\{\left(\frac{\partial \ln L}{\partial\sigma_1^2}\right)^2\right\}=-E\left(\frac{\partial^2 \ln L}{\partial\sigma_1^2\partial\sigma_1^2}\right)\left\{\frac{1-c_k+(T+1)a_k^2/2}{(1-2a_k)}\right\}$$

where $c_k=(3b_k+2a_k)/2$. Now, for all values of $T$ that are ever likely to be relevant ($<1000$ say) the term $(T+1)a_k^2$ is at most $10^{-11}(k\geqq 6)$ so that it seems reasonable to assume (and the Monte Carlo results which follow confirm this) that as $T$ becomes large in a practical sense, the distribution of $T_1$ will be nearly normal. However, as $T\to\infty$, $\surd(T/2)((\hat{\sigma}_1^2/\sigma_1^2)-1)$ will not tend to normality and to avoid this a correction can be made by considering

$$\sqrt{\frac{T}{2}}\left(\frac{\hat{\sigma}_1^2}{\sigma_1^2}-\frac{1-3a_k}{1-2a_k}\right)\frac{(1-2a_k)}{\left(1-\frac{a_k(4+k)^2}{2}\right)^{\frac{1}{2}}}$$

and for all practical purposes, this correction is small enough to ignore.

To derive $\eta$, we differentiate (30) w.r.t. $\theta_{1j}$ in turn. Thus, for $j=1,\ldots,k_1$

$$\eta_j=\frac{T}{2\sigma_{01}^2}\frac{\partial\sigma_{01}^2}{\partial\theta_{1j}}+\sum_t\frac{\partial E_1(\ln y_t)}{\partial\theta_{1j}}.$$

But from (32), using the exact rather than the estimated form,

$$\frac{\partial\sigma_{01}^2}{\partial\theta_{1j}}=\frac{2}{T}\sum_t\left[\text{cov}_1\left\{\ln y_t,\, y_t^{-1}\frac{\partial g_t(\theta_1)}{\partial\theta_{1j}}\right\}+\{E_1(\ln y_t)-f_t(\theta_{01})\}E_1\left\{y_t^{-1}\frac{\partial g_t(\theta_1)}{\partial\theta_{1j}}\right\}\right]$$

since, from (33), $\sum_t\{E_1(\ln y_t)-f_t(\theta_{01})\}\partial f_t(\theta_{01})/\partial\theta_{01j}=0,\quad\forall_j$. Hence, for $j=1,\ldots,k_1$, on simplification

$$\eta_j=\sum_t\frac{\partial g_t(\theta_1)}{\partial\theta_{1j}}E_1\left\{\frac{\ln y_t-f_t(\theta_{01})}{\sigma_{01}^2 y_t}\right\}+\sum_t\frac{\partial g_t(\theta_1)}{\partial\theta_{1j}}E_1\left\{\frac{1}{y_t}\right\}\qquad\ldots(36)$$

Similarly, differentiating (30) w.r.t. $\sigma_1^2$,

$$\eta_{k_1+1}=\frac{T}{2}\left[\frac{1}{\sigma_{01}^2}\frac{\partial\sigma_{01}^2}{\partial\sigma_1^2}-\frac{1}{\sigma_1^2}\right]+\sum_t\frac{\partial E_1(\ln y_t)}{\partial\sigma_1^2}.$$

It may easily be checked from the distribution of $\varepsilon_{1t}$ that

$$\frac{\partial E_1\{\ln y_t\}}{\partial\sigma_1^2}=\frac{1}{2\sigma_1^2}E_1\left\{\frac{\varepsilon_{1t}}{y_t}\right\}.$$

Hence,

$$\frac{\partial\sigma_{01}^2}{\partial\sigma_1^2}=\frac{1}{T}\sum_t E_1\left[\frac{\varepsilon_{1t}}{\sigma_1^2 y_t}\{\ln y_t-f_t(\theta_{01})\}\right].$$

Thus, collecting terms

$$\eta_{k_1+1}=-\frac{T}{2\sigma_1^2}+\frac{1}{2\sigma_1^2}\sum_t E_1\left[\frac{\varepsilon_{1t}\{\ln y_t-f_t(\theta_{01})\}}{\sigma_{01}^2 y_t}\right]+\frac{1}{2\sigma_1^2}\sum_t E_1\left\{\frac{\varepsilon_{1t}}{y_t}\right\}.\qquad\ldots(37)$$

Finally, from (10) and substitution of $\alpha_{01}$ in (9), $L_{01}=L_1(\alpha_1)-L_0(\alpha_{01})$ is given by

$$L_{01}=\frac{T}{2}\ln\left(\frac{\sigma_{01}^2}{\sigma_1^2}\right)-\frac{\sum_t \varepsilon_{1t}^2}{2\sigma_1^2}+\sum_t \ln y_t+\frac{1}{2\sigma_{01}^2}\sum_t\{\ln y_t-f_t(\theta_{01})\}^2+T\ln\alpha(k). \quad \ldots(38)$$

From this $V_1(L_{01})$ can be calculated term by term. Since very few of the relevant expectations can be simplified analytically, there is little point in deriving further algebraic expressions for this variance and the questions of computation will be taken up below. Once $V_1(L_{01})$ has been calculated, $V_1(T_1)$ is calculated from, using (34),

$$V_1(T_1)=V_1(L_{01})-\zeta' M_g^{-1}\zeta\frac{\sigma_1^2}{T}-\frac{2\sigma_1^4}{T}\eta_{k_1+1}^2(1-2a_k)^{-1} \quad \ldots(39)$$

using (36) and (37) and replacing all unknown quantities by their ML estimates. (The quantity $a_k$ can be taken to be zero in practice.)

## 2. THE COMPUTATIONS OF THE TEST STATISTICS

The computation of $T_0$ and $V_0(T_0)$ poses no special difficulty. In general, both $H_0$ and $H_1$ will be estimated by non-linear techniques and the calculation of $T_0$ requires one additional non-linear regression to estimate $\hat{\theta}_{10}$ and $\hat{\sigma}_{10}^2$. The expression for $V_0(T_0)$, (29), although complicated, is trivial enough to calculate. Unfortunately, this is not the case for $T_1$ or $V_1(T_1)$. For example, we require the expression $E_1(\ln y_t)$ which is given by

$$E_1(\ln y_t)=\int_{-k\sigma_1}^{k\sigma_1}\frac{\ln\{g_t(z,\theta_1)+\tau\}}{\sqrt{2\pi}\sigma_1}\exp\left\{-\tfrac{1}{2}\frac{\tau^2}{\sigma_1^2}\right\}d\tau, \quad \ldots(40)$$

for some suitable value of $k$, as well as a range of other expectations which arise in the calculation of $V_1(T_1)$. Expressions like (40) can be evaluated each time they arise by numerical integration procedures but, in practice, this is prohibitively expensive even for practical work (e.g. if $T=60$, there would be 600 or so integrations required for one test statistic), let alone for Monte Carlo experiments.

We therefore adopt the following alternative. Define $\upsilon$, a truncated $N(0,1)$ variable, by

$$\upsilon=\varepsilon_1/\sigma_1. \quad \ldots(41)$$

We can write $\pi(\upsilon)$ for its density function so that (40) becomes

$$E_1(\ln y_t)=\int_{-k}^{k}\{\ln g_t(\theta_1)+\ln(1+\xi\upsilon)\}\pi(\upsilon)d\upsilon$$

$$=\ln g_t(\theta_1)+\int_{-k}^{k}\ln(1+\xi\upsilon)\pi(\upsilon)d\upsilon, \quad \ldots(42)$$

where $\xi=\sigma_1/g_t(\theta_1)$. On working through the expressions for $T_1$ and its variance, all required expectations can be evaluated with the aid of integrals of the form

$$I(a,b,c)=\int_{-k}^{k}(1+\xi\upsilon)^a[\ln(1+\xi\upsilon)]^b\upsilon^c\pi(\upsilon)d\upsilon, \quad \ldots(43)$$

where $\xi$ and $\upsilon$ are as defined above and we require the ten integrals $I(0,1,0)$, $I(0,2,0)$, $I(0,3,0)$, $I(0,4,0)$, $I(0,1,2)$, $I(0,2,2,)$, $I(-1,1,0)$, $I(-1,1,1)$, $I(-1,0,1)$, and $I(-1,0,0)$. For given $k$, these integrals define ten functions of $\xi$, which we require for all values of $\xi$ between zero and, say, 0·125. (If $g_t(\theta_1)\not> 8\sigma_1$, it is not reasonable to consider $H_0$ as a serious possibility.)

Each of the integrals was thus evaluated by numerical integration for values of $\xi$ from 0 to $\frac{1}{8}$ by intervals of 1/1024; the calculated values were then approximated by Chebycheff polynomials. The polynomial coefficients were then built into the computer programme and used in routine evaluations of the statistics. This procedure worked extremely well. The calculations were very rapid compared with repeated quadrature and we were able to show that for reasonable values of $k$ (6 to 10), the integrals were not sensitive to the precise value chosen. The programme was written so as to reject any value of $\xi > 0{\cdot}125$; in this case, the Chebycheff approximations become unreliable but this is of no importance since, if this happens, $H_0$ is clearly incorrect *a priori.*

## 3. SOME EMPIRICAL EVIDENCE

In this section, we present empirical evidence designed to elucidate the properties of the test, in particular, to investigate its distribution when either $H_0$ or $H_1$ is true and to discover how this is affected by sample size, and by other considerations. We also compare the Cox test with the unmodified likelihood ratio criterion suggested by Sargan (1964). Finally, we offer some tentative evidence for the case when neither $H_0$ nor $H_1$ is correct as specified. This last is perhaps the most important potential application of the test, but the possibilities are too vast to be more than touched on here.

Of particular interest for applications of the test is the question of how much guidance the large sample distributions of the statistics gives us about their behaviour in small or moderate-sized samples. The theory tells us that $N_0 = T_0/\sqrt{V_0}$ is asymptotically distributed as $N(0, 1)$ under $H_0$ and similarly for $N_1$ under $H_1$. However, we are ultimately interested in the joint distribution of $N_0$ and $N_1$ under both hypotheses and since both statistics are functions of the same magnitude, the log likelihood ratio, this can be done straightforwardly, at least asymptotically. To clarify, we change the notation slightly and write, from (5)

$$N_0 = V_0(\hat{\alpha}_0)^{-\frac{1}{2}}\{\hat{L}_{10} - P_0(\hat{\alpha}_0)\} \qquad \ldots(44)$$

$$N_1 = V_1(\hat{\alpha}_1)^{-\frac{1}{2}}\{-\hat{L}_{10} + P_1(\hat{\alpha}_1)\} \qquad \ldots(45)$$

where $P_0(\hat{\alpha}_0) = T\{\text{plim}_0\,(\hat{L}_{10}/T)\}_{\alpha_0=\hat{\alpha}_0}$, $P_1(\hat{\alpha}_1) = T\{\text{plim}_1\,(\hat{L}_{10}/T)\}_{\alpha_1=\hat{\alpha}_1}$, we have used $\hat{L}_{10} = -\hat{L}_{01}$, and the notation $V_0(\ )$, $V_1(\ )$, $P_0(\ )$ and $P_1(\ )$, emphasizes that the various magnitudes are evaluated as functions of the parameters of the current working hypothesis taken at their maximum likelihood values. For example, $V_0$ is given by (29) as a function of $\alpha_0$ and $\alpha_{10}$ but the latter is itself a function of $\alpha_0$ via (23) and (24) and, in practice, (29) and thus $N_0$ is evaluated using the ML estimate of $\alpha_0$. If we rearrange (44) and (45), $N_0$ and $N_1$ satisfy (exactly),

$$N_0 V_0(\hat{\alpha}_0)^{\frac{1}{2}} + N_1 V_1(\hat{\alpha}_1)^{\frac{1}{2}} = P_1(\hat{\alpha}_1) - P_0(\hat{\alpha}_0). \qquad \ldots(46)$$

Hence, under $H_0$

$$\left[\frac{V_1(\alpha_{10})}{V_0(\alpha_0)}\right]^{\frac{1}{2}}\left[N_1 - \frac{\{P_1(\alpha_{10}) - P_0(\alpha_0)\}}{V_1(\alpha_{10})^{\frac{1}{2}}}\right] \underset{a}{\sim} N(0, 1) \qquad \ldots(47)$$

while under $H_1$

$$\left[\frac{V_0(\alpha_{01})}{V_1(\alpha_1)}\right]^{\frac{1}{2}}\left[N_0 - \frac{\{P_1(\alpha_1) - P_0(\alpha_{01})\}}{V_0(\alpha_{01})^{\frac{1}{2}}}\right] \underset{a}{\sim} N(0, 1). \qquad \ldots(48)$$

The situation when $H_0$ is true is illustrated in Figure 1; as the sample becomes large $N_0$ and $N_1$ will lie along the line ABCD. A similar situation pertains when $H_1$ is true although the line will have a different position and slope. In both cases ABCD moves away from the origin, maintaining its slope, as the sample size increases. The joint distribution of $N_0$ and $N_1$ is thus a singular one above the line ABCD. If $H_0$ is true its mean lies always above the

$N_0$ axis, if $H_1$ is true above the $N_1$ axis, while if neither is true we have no information about the shape other than that the distribution lies along the line.

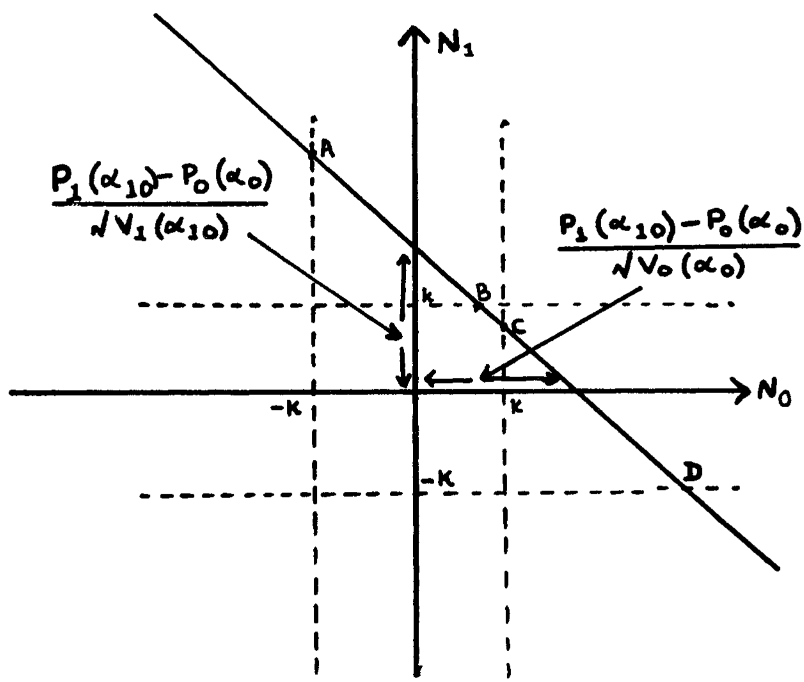


FIGURE 1
Asymptotic joint distribution of $N_0$ and $N_1$ under $H_0$

In Monte Carlo experiments, when the parameters of the true models are known, the quantities in (47) can be calculated when $H_0$ is true and those in (48) when $H_1$ is true. We can thus calculate the asymptotic theoretical rejection and acceptance frequencies and compare these with the empirical results. Indeed, had we been fully aware of the technique earlier, the Monte Carlo experiments could have been designed around these asymptotic formulae, see Mizon and Hendry (1980). As it is, our attempts to use the asymptotic formulae have not so far been successful and for the present we content ourselves with the discussion of conventional Monte Carlo results.

In order to keep the analysis as simple as possible and so as to keep the computations within bounds, the experiments are designed around the simplest linear model

$$y_t = \alpha + \beta x_t + u_t, \qquad \ldots(49)$$

where $x_t$ is an autoregressive process defined by

$$(x_t - \mu) = \rho(x_{t-1} - \mu) + \varepsilon_t, \qquad \ldots(50)$$

with $u_t \sim N(0, \sigma_u^2)$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. The expectation of $x_t$, $\mu$, is chosen so that the series is positive; hence, (49) can be compared with the loglinear regression equation

$$\ln y_t = \alpha^* + \beta^* \ln x_t + u_t^*. \qquad \ldots(51)$$

Working first with the case where the linear model, $H_1$, as given by (49) is true, our ability to discriminate the true model from the false $H_0$, (51), is likely to depend upon a number of factors. First, the choice of $\alpha$ and $\beta$ is important. If $\alpha$ is very large relative to $\beta\mu$, and ignoring the effects of the error, then (51) will approximate (49) rather well with

$\alpha^* = \ln \alpha$. Similarly, if $\alpha$ is close to zero, setting $\alpha^* = \ln \beta$ and $\beta^* = 1$ will make (51) close to (49). Consequently, values of $\alpha$ close to $\beta\mu$ are likely to give us the best chance of discrimination. The second important factor is the variance of $x_t$; hence $\rho$ and $\sigma_\varepsilon$ are likely to be important. Finally, discrimination is likely to depend on $\sigma_u$; when this is small, the linear model should fit so well as to rule out alternative specifications.

The experiments were carried out as follows. At a preliminary stage, six $x_t$ series, each of 80 observations, were generated corresponding to $\rho = (0{\cdot}5,\ 0{\cdot}7,\ 0{\cdot}9)$ and $\sigma_\varepsilon = (8, 16)$, using the normal random number generator GØ5ADF from the NAGLIB library. In each case, sufficient early values were discarded so as to ensure the arbitrary initial choice for $x_0$ had negligible importance. The parameter $\mu$ was set to 100 in all cases. All generated values of $x_t$ were positive. The parameter $\sigma_u^2$ was controlled indirectly by setting the asymptotic $R^2$ for the regression equation (49). Elementary manipulation leads to

$$\text{plim } R^2 = \frac{\beta^2\sigma_\varepsilon^2/(1-\rho^2)}{\sigma_u^2 + \beta^2\sigma_\varepsilon^2/(1-\rho^2)}. \qquad \ldots(52)$$

Given the extremely small probability of generating values of $u_t$ further than $8\sigma_u$ from zero, no attempt was made to truncate the distribution. Given $\sigma_\varepsilon^2$ and $\rho$, $\sigma_u^2$ can be chosen to set $R^2$ at the desired level using (52). Three different sample sizes were used; for each replication, 80 observations of $y_t$ were generated corresponding to $x_t$. These were first used in a single test with sample size 80; each half of the sample ($y_t$ and $x_t$) was then used for two more tests of sample size 40. Finally, each half was itself subdivided to give four tests each with sample size 20. Each complete experiment was replicated 500 times. There are thus 500 replications for sample sizes 80, 1000 for sample sizes 40, and 2000 for sample size 20. The value of $\beta$ was set to 5 throughout, with $\alpha$ at 500 ($=\beta\mu$). Due to an oversight, it was not noticed that it is impossible to vary $R^2$ for these values of $\alpha$ and $\beta$ without creating a high probability of negative values of $y_t$. Thus, in the experiments where $R^2$ is varied, $\alpha$ is set to 1000. The differences in results between $\alpha = 1000$ and $\alpha = 500$ were small enough to justify not repeating the earlier experiments.

Before discussing the results, we describe briefly the layout of the tables where it is not self-explanatory. In Tables I to III the first column relates to the parameter being varied, while the second, $T$, give the relevant sample size, 80, 40, or 20. The four columns headed by $\pi_0$ through to $\pi_{10}$ give the fractions of all cases in which the hypothesis subscripted was *rejected*. Hence, in the first row of Table I, 14·8 per cent of the cases rejected $H_0$, the loglinear model, 78·6 per cent of the cases rejected neither, 2·4 per cent rejected $H_1$, the linear model, and 4·2 per cent rejected both. The usual 5 per cent points on the normal distribution were used. Hence, the $\pi$ figures were constructed to the rules

($\pi_0$) Reject $H_0$ and accept $H_1$ whenever $N_0 = T_0/\sqrt{V_0}$ and $N_1 = T_1/\sqrt{V_1}$ are such that $|N_0| \geqq 1{\cdot}96$ and $|N_1| < 1{\cdot}96$

($\pi_1$) Reject $H_1$ and accept $H_0$ whenever $|N_0| < 1{\cdot}96$ and $|N_1| \geqq 1{\cdot}96$

($\pi$) Accept both $H_0$ and $H_1$ whenever $|N_0| < 1{\cdot}96$ and $|N_1| < 1{\cdot}96$

($\pi_{10}$) Reject both $H_0$ and $H_1$ whenever $|N_0| \geqq 1{\cdot}96$ and $|N_1| \geqq 1{\cdot}96$.

These four probabilities are illustrated in Figure 1 for $H_0$ true and for a critical value of $k$ ($=1{\cdot}96$ in this case). $\pi_{10}$ corresponds to that part of the line to the north-west of A and the south-east of D, $\pi_0$ to AB, $\pi$ to BC, and $\pi_1$ to CD. Note that as the sample size rises, and ABCD moves outwards keeping its slope constant, $\pi$ will eventually fall to zero and the power $(\pi_1 + \pi_{10})$ to unity.

Column 7, headed $S_0$, gives the proportion rejecting $H_0$, i.e. favouring $H_1$, on the Sargan likelihood ratio criterion. Clearly, no significance level can be set for a "test" such as this and we have followed what would presumably be practice using the method,

TABLE I

*Linear model ($H_1$) is true*

$\rho = 0{\cdot}9$, $R^2 = 0{\cdot}9$, $\alpha = 500$

| $\sigma_\varepsilon$ | $T$ | $\pi_0$ | $\pi$ | $\pi_1$ | $\pi_{10}$ | $S_0$ | $\bar{N}_0 : \hat{V}_0$ | $\bar{N}_1 : \hat{V}_1$ | $KS_1$ |
|---|---|---|---|---|---|---|---|---|---|
| 8 | 80 | 0·148 | 0·786 | 0·024 | 0·042 | 0·708 | −1·09 | −0·00 | 0·0247 |
| | | (0·016) | (0·018) | (0·007) | (0·009) | (0·020) | 1·03 | 1·09 | |
| 8 | 40 | 0·071 | 0·876 | 0·021 | 0·032 | 0·588 | −0·57 | −0·05 | 0·0342 |
| | | (0·008) | (0·010) | (0·005) | (0·006) | (0·016) | 1·09 | 1·02 | |
| 8 | 20 | 0·027 | 0·931 | 0·015 | 0·027 | 0·525 | −0·26 | −0·07 | 0·0443** |
| | | (0·004) | (0·006) | (0·003) | (0·004) | (0·011) | 0·95 | 0·93 | |
| 16 | 80 | 0·574 | 0·362 | 0·038 | 0·026 | 0·828 | −2·22 | −0·03 | 0·0542 |
| | | (0·022) | (0·022) | (0·009) | (0·007) | (0·017) | 1·15 | 1·15 | |
| 16 | 40 | 0·241 | 0·708 | 0·023 | 0·028 | 0·700 | −1·24 | −0·07 | 0·0472* |
| | | (0·014) | (0·014) | (0·005) | (0·005) | (0·015) | 1·28 | 1·28 | |
| 16 | 20 | 0·080 | 0·884 | 0·014 | 0·022 | 0·590 | −0·59 | −0·10 | 0·0717** |
| | | (0·006) | (0·007) | (0·003) | (0·003) | (0·011) | 1·01 | 1·01 | |

TABLE II

*Linear model ($H_1$) is true*

$\rho = 0{\cdot}7$, $\sigma_\varepsilon = 16$, $\alpha = 1000$

| $R^2$ | $T$ | $\pi_0$ | $\pi$ | $\pi_1$ | $\pi_{10}$ | $S_0$ | $\bar{N}_0 : \hat{V}_0$ | $\bar{N}_1 : \hat{V}_1$ | $KS_1$ |
|---|---|---|---|---|---|---|---|---|---|
| 0·9 | 80 | 0·708 | 0·238 | 0·032 | 0·022 | 0·884 | −2·70 | −0·08 | 0·0557 |
| | | (0·020) | (0·019) | (0·008) | (0·007) | (0·014) | 1·31 | 1·04 | |
| 0·9 | 40 | 0·357 | 0·593 | 0·036 | 0·014 | 0·733 | −1·51 | −0·10 | 0·0483* |
| | | (0·015) | (0·016) | (0·006) | (0·004) | (0·014) | 1·89 | 1·02 | |
| 0·9 | 20 | 0·233 | 0·709 | 0·041 | 0·017 | 0·656 | −1·06 | −0·12 | 0·0470** |
| | | (0·009) | (0·010) | (0·004) | (0·003) | (0·011) | 1·61 | 1·08 | |
| 0·7 | 80 | 0·348 | 0·592 | 0·034 | 0·026 | 0·758 | −1·61 | −0·06 | 0·0405 |
| | | (0·021) | (0·022) | (0·008) | (0·007) | (0·019) | 1·27 | 1·09 | |
| 0·7 | 40 | 0·162 | 0·780 | 0·027 | 0·031 | 0·668 | −0·96 | −0·07 | 0·0462* |
| | | (0·012) | (0·013) | (0·005) | (0·006) | (0·015) | 1·29 | 1·06 | |
| 0·7 | 20 | 0·101 | 0·844 | 0·032 | 0·023 | 0·606 | −0·65 | −0·09 | 0·0403** |
| | | (0·007) | (0·008) | (0·004) | (0·003) | (0·011) | 1·18 | 1·03 | |

TABLE III

*Loglinear model ($H_0$) is true*

$\rho = 0{\cdot}9$, $R^2 = 0{\cdot}9$

| $\sigma_\varepsilon$ | $T$ | $\pi_0$ | $\pi$ | $\pi_1$ | $\pi_{10}$ | $S_1$ | $\bar{N}_0 : \hat{V}_0$ | $\bar{N}_1 : \hat{V}_1$ | $KS_0$ |
|---|---|---|---|---|---|---|---|---|---|
| 0·08 | 80 | 0·018 | 0·782 | 0·176 | 0·024 | 0·694 | −1·09 | −1·07 | 0·0594 |
| | | (0·006) | (0·019) | (0·017) | (0·007) | (0·021) | 0·98 | 1·11 | |
| 0·08 | 40 | 0·021 | 0·897 | 0·056 | 0·026 | 0·614 | −0·04 | −0·58 | 0·0359 |
| | | (0·005) | (0·010) | (0·007) | (0·005) | (0·015) | 0·95 | 0·95 | |
| 0·08 | 20 | 0·010 | 0·942 | 0·026 | 0·023 | 0·553 | −0·03 | −0·30 | 0·0299 |
| | | (0·002) | (0·005) | (0·004) | (0·003) | (0·011) | 0·89 | 0·89 | |
| 0·16 | 80 | 0·024 | 0·392 | 0·562 | 0·022 | 0·828 | −0·15 | −2·23 | 0·0826** |
| | | (0·007) | (0·022) | (0·022) | (0·007) | (0·017) | 1·00 | 1·32 | |
| 0·16 | 40 | 0·026 | 0·757 | 0·192 | 0·025 | 0·696 | −0·09 | −1·13 | 0·0437* |
| | | (0·005) | (0·014) | (0·013) | (0·005) | (0·015) | 1·00 | 1·09 | |
| 0·16 | 20 | 0·015 | 0·913 | 0·055 | 0·018 | 0·604 | −0·10 | −0·57 | 0·0568** |
| | | (0·003) | (0·006) | (0·005) | (0·003) | (0·011) | 0·87 | 0·87 | |

to accept the model with the higher likelihood. Columns 8 and 9 headed $\bar{N}_0$: $\hat{V}_0$ and $\bar{N}_1$: $\hat{V}_1$ give the sample means and variances of the $N_0$'s and $N_1$'s actually calculated; in theory if $H_0$ is true, $\bar{N}_0$ should be zero and $\hat{V}_0$ be unity, similarly for $\bar{N}_1$ and $\hat{V}_1$. The final column $KS_1$ or $KS_0$ is the value of the Kolmogorov–Smirnov test of the hypothesis that the $N_1$'s or $N_0$'s are distributed as $N(0, 1)$. Significant departures from $N(0, 1)$ are indicated by * at 5 per cent and ** at 1 per cent. All figures in brackets are estimated standard errors.

We have not presented our results in terms of the usual Type I and Type II errors because the use of these concepts is less attractive when there are four rather than two possible decisions. However, if required, the probability of Type I error is given by $\pi_1 + \pi_{10}$ in Tables I and II and by $\pi_0 + \pi_{10}$ in Table III, while the probabilities of Type II error are $\pi + \pi_1$ and $\pi + \pi_0$ respectively.

The first set of experiments take $\alpha = \beta\mu = 500$, $R^2 = \rho = 0{\cdot}9$ with $\sigma_\varepsilon = (8, 16)$; the results are given in Table I. These were repeated with $\rho = 0{\cdot}7$ and $0{\cdot}5$ for the same settings of the other parameters but the results were very close to those for $\rho = 0{\cdot}9$ and are not given here. Table II investigates the effects of changing $R^2$ from 0·9 to 0·7 with $\rho = 0{\cdot}7$ and $\alpha = 1000$.

Looking at the last column first, the Kolmogorov–Smirnov statistics show no evidence that the Cox statistics are not distributed as $N(0, 1)$, for sample size 80. For sample size 40, 3 out of the 4 cases indicate a departure at the 5 per cent level, although in the four experiments not reported in detail ($\rho = 0{\cdot}5$, $0{\cdot}7$ with $\sigma_\varepsilon = 8, 16$), all but one are consistent with $N(0, 1)$. However, when $T = 20$, all four cases shown indicate rejection at the 1 per cent level. Note however that the Kolmogorov–Smirnov test is extremely powerful and applies to the whole distribution. There is no evidence at all in these results to suggest that the probability of Type I error is significantly larger than 0·05, even when $T = 20$. This is particularly important since it is frequently conjectured that large sample tests, such as the Cox test, are prone to over-frequent rejection of correct hypotheses when used in small sample situations.

Given the Kolmogorov–Smirnov test results, the values for $\bar{N}_1$ and $\hat{V}_1$ are as expected being, in all cases, close to 0 and 1 respectively. The values of $\bar{N}_0$ and $\hat{V}_0$ illustrate that the test performs qualitatively as it should; when the false hypothesis is treated as if it were true, the true model fits better than one would expect it to. When $H_0$ is false, the distribution of $N_0$ is shifted to the left to an extent which increases with the sample size and with the noisiness of the independent variable ($\sigma_\varepsilon$), while decreasing with the noisiness of the true model ($\sigma_u$). Note that, in these experiments at least, the variance of $N_0$ when $H_1$ is true remains, in most cases, close to unity.

These shifts in $N_0$ determine the performance of the test, the relevant characteristics of which are summarized in columns 3–6. The correct decision is to reject $H_0$ and the frequency this occurs varies from 71 per cent to 3 per cent depending on $R^2$, $\sigma_\varepsilon$ and $T$. Since columns 5 and 6 add to approximately 0·05 (the Type I error), column 4, the no decision case, is approximately equal to 0·95 less column 3. In other words, apart from the constant 5 per cent error, the test either makes the correct decision or is indecisive, the proportion of one to the other being determined by $R^2$, $\sigma_\varepsilon$ and $T$. The Sargan test, $S_0$, has no possibility of indecision, simply selecting the model with the higher likelihood. Not surprisingly, the fraction of successes, $S_0$, responds much as does $\pi_0$ to changes in $R^2$, $\sigma_\varepsilon$ and $T$; further, it is always greater than both $\pi_0$ and 0·5 (its expectation given zero discriminatory power). Thus the Sargan test contains useful information on choosing between the two models and, given its extreme simplicity of calculation–it requires no more than the original estimation–it is likely to be useful in practice, *provided we are certain in advance that either $H_0$ or $H_1$ is true.* The superiority of $S_0$ over the Cox procedure is due to its "one or other" nature; there is no possibility of indecision nor is it possible for both models to be rejected. We shall see what happens when $H_0$ and $H_1$ are both false below. Note too that the two models being compared here have identical numbers of parameters so that the tendency of pure likelihood tests to favour the model with the greater number of parameters is of no consequence. Without some correction however, the Sargan test is likely to be dangerous unless the number of parameters are the same in both models, i.e. in the case of a "pure" log versus linear comparison. More generally some correction to favour more "parsimonious" models could easily be built in, for example by taking $\hat{L}_{10}-(k_0-k_1)$ rather than $\hat{L}_{10}$ in which case the Sargan test is equivalent to using the Akaike ((1972) and (1974)) information criterion. (See also Sawa (1978) for further discussion.)

Table III provides a check that the test works both ways round and should be read in conjunction with Table I. In this case $H_0$ is true, so that the $\pi_1$ column corresponds to the correct decision and thus to $\pi_0$ in Table I. The true model underlying these results is

$$\ln y_t = \gamma + \delta \ln x_t + v_t \qquad \text{...(53)}$$

and

$$\ln x_t - \ln \mu^* = \rho(\ln x_{t-1} - \ln \mu^*) + \varepsilon_t, \qquad \text{...(54)}$$

where $\gamma = 4{\cdot}6$, $\delta = 0{\cdot}5$, $\mu^* = 100$, $\sigma_\varepsilon = (0{\cdot}08,\ 0{\cdot}16)$, $\rho = 0{\cdot}9$, all these numbers being chosen so as to make (53) a close approximation to the original linear model (49). This device seems to have been successful in that Table III replicates very closely the corresponding numbers in Table I. Note, however, that when $\sigma_\varepsilon = 0{\cdot}16$, the Kolmogorov–Smirnov test rejects $N(0, 1)$ for all three sample sizes, the only case where this occurs in all the experiments undertaken.

Finally, we look briefly at the case where both $H_0$ and $H_1$ are false so that we are using the Cox test as a test of misspecification. There is no particular reason to expect the test to be generally powerful in this context; it is designed around $H_0$ and $H_1$ specifically and its performance in recognising misspecification is likely to depend very much on the alternative considered. We look at only one example, albeit one which is likely to arise quite often in practice. The data were generated according to

$$y_t = \alpha + \beta x_t + z_t + u_t, \qquad \text{...(55)}$$

where $\alpha$, $\beta$, $x_t$ and $u_t$ are as in (49) and (50), with $z_t$, an independent autoregressive process given by

$$z_t = \rho_2 z_{t-1} + \varepsilon_{2t}, \qquad \text{...(56)}$$

where $\varepsilon_{2t} \sim N(0, \sigma^2_{2\varepsilon})$. $\rho_2$ was set at $0{\cdot}7$ and $\sigma_2^2$ varied over (10, 20, 30). $H_0$ and $H_1$ were as before; hence, $H_1$ is correct but for an omitted variable $z_t$, the importance of which varies with $\sigma_{2\varepsilon}$, whereas $H_0$ is incorrect, not only in omitting a variable, but also in functional form. $H_1$ is thus likely to be a better approximation to (55) than is $H_0$. The variance of $u_t$, $\sigma^2_u$, was set so as to give an asymptotic $R^2$ of $0{\cdot}9$ on the misspecified equation $H_1$; it can easily be checked that this is possible for the values of $\alpha$, $\beta$, $\rho_2$ and $\sigma_{2\varepsilon}$ indicated. It is important to realise that, in the results which follow, the (incorrect) equation $H_1$ fits as well as it did when it was correct in the earlier experiments.

The results of the tests for $T = 80$ and $T = 40$ are given in Table IV. Experiments for $T = 20$ are not presented since these, in the vast majority of cases, gave a no decision result. Clearly, the sample size is very important in these experiments, as is the variability of the omitted variable. In all experiments, $H_0$ is rejected much more frequently than is $H_1$. The Sargan test, too, always favour $H_1$ by a large majority. When $T = 40$, most experiments lead to no result with most of the remainder rejecting $H_0$ only. However, when the sample size increases to 80, the test begins to reject both models, 6% of the time when $\sigma_{2\varepsilon} = 10$, $22{\cdot}4$ per cent of the time when $\sigma_{2\varepsilon} = 20$ and $78{\cdot}6$ per cent of the time when $\sigma_{2\varepsilon} = 30$. Note that, in this case, the likelihood ratio favours $H_1$ in all of the 500 cases so that the Sargan test, although decisive, is decisively wrong. Parenthetically, it is worth noting that the shifts in the distribution of $N_0$ and $N_1$ now appear to be much more complicated than when one of $H_0$ and $H_1$ was true. The negative values of $N_0$ reflect the

TABLE IV

*Neither model is true*

| $\sigma_{2\varepsilon}$ | $T$ | $\pi_0$ | $\pi$ | $\pi_1$ | $\pi_{10}$ | $S_0$ | $\bar{N}_0 : \hat{V}_0$ | $\bar{N}_1 : \hat{V}_1$ |
|---|---|---|---|---|---|---|---|---|
| 10 | 80 | 0·562 | 0·374 | 0·004 | 0·060 | 0·920 | −2·24 | 0·62 |
| | | (0·022) | (0·022) | (0·003) | (0·011) | (0·012) | 0·95 | 0·92 |
| 10 | 40 | 0·164 | 0·784 | 0·017 | 0·035 | 0·676 | −0·94 | 0·06 |
| | | (0·012) | (0·013) | (0·004) | (0·006) | (0·015) | 1·41 | 1·03 |
| 20 | 80 | 0·652 | 0·124 | 0·000 | 0·224 | 0·986 | −2·95 | 1·32 |
| | | (0·021) | (0·015) | (0·000) | (0·019) | (0·005) | 0·71 | 0·65 |
| 20 | 40 | 0·206 | 0·734 | 0·011 | 0·049 | 0·697 | −1·11 | 0·18 |
| | | (0·013) | (0·014) | (0·003) | (0·007) | (0·015) | 1·67 | 1·02 |
| 30 | 80 | 0·214 | 0·000 | 0·000 | 0·786 | 1·000 | −3·96 | 2·25 |
| | | (0·018) | (0·000) | (0·000) | (0·018) | (0·000) | 0·19 | 0·15 |
| 30 | 40 | 0·397 | 0·534 | 0·000 | 0·069 | 0·673 | −1·59 | 0·46 |
| | | (0·016) | (0·016) | (0·000) | (0·008) | (0·015) | 2·72 | 1·08 |

preference for $H_1$ while the positive values of $N_1$ suggest that $H_1$ is false but should not be replaced by $H_0$. However, these positive and negative values arise as $T$ increases by a "splitting" rather than a shifting of the distributions. When $T$ is 20, both distributions are heavily concentrated in the no decision area. When $T$ is 40, both are bi-modal with peaks in no decision and reject zones for $H_0$ and with two peaks in the no decision zone for $H_1$. At $T = 80$, both central peaks vanish and the outer peaks shift outward in opposite directions giving the result in the second last row of the table.

Clearly, this is only one example of misspecification but we believe these results are encouraging enough to suggest that further work on specification analysis using the Cox test would be well worth attempting.

## 4. SUMMARY AND CONCLUSIONS

In this paper, we have applied Cox's procedure for non-tested hypotheses tests to the problem of testing a logarithmic versus a linear model. Section 1 derived the statistics and presented formulae for their calculation. We also presented empirical evidence which suggests that the large sample properties of the test are sufficiently closely realised in quite small samples to make the test practical. We found no evidence, in even very small samples, of a tendency for the test to reject correct hypotheses too frequently. The power of the test, however, depended crucially on the sample size, on the fit of the true hypothesis and on the noisiness of the independent variable. Our results also suggested that, provided the investigator knows that one or other of the models formally considered is correct, the unmodified likelihood ratio, as suggested by Sargan, is a useful discriminator between the models, at least when both have the same number of parameters. In the case where neither model is true, we presented some evidence that, with sufficiently large samples, the Cox test can detect misspecification even when both models fit well according to conventional criteria.

Clearly, there is great scope for further research. The basic results of Tables I to III need to be confirmed on more general models, particularly in cases where non-linear estimation is involved. Perhaps most exciting, however, are applications to other types of misspecification analysis involving choice of functional form as well as omitted variables. Such tests should also include an examination of the role of the normality assumption particulary given Amemiya's (1977) recent work on the importance of normality in the maximum likelihood estimation of the general non-linear model. By extension of Amemiya's results, it may turn out that while the Cox test is robust against non-normality when errors are additive, the robustness may not extend to the cases examined in this paper.

We should like to thank David Hendry, Adrian Pagan, Peter Phillips, Jean-Francois Richard, Gene Savin and especially Hashem Pesaran for extremely helpful comments. David Mitchell gave invaluable assistance with the computations.

REFERENCES

AKAIKE, H. (1972), "Information theory and an extension of the maximum likelihood principle", in *Proc. 2nd Int. Symp. on Information Theory*, 267–281.

AKAIKE, H. (1974), "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, Ac-19, 716–723.

AMEMIYA, T. (1973), "Regression analysis when the dependent variable is truncated normal", *Econometrica*, **41**, 997–1016.

AMEMIYA, T. (1976), "Selection of regressors" (Stanford University Institute for Mathematical Studies in the Social Sciences, Technical Report No. 225).

AMEMIYA, T. (1977), "The maximum likelihood and the nonlinear three-stage least squares estimator in the general non-linear simultaneous equation model", *Econometrica*, **45**, 955–968.

COX, D. R. (1961), "Tests of separate families of hypotheses" *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**, (Berkeley: University of California Press).

COX, D. R. (1962), "Further results on tests of separate families of hypotheses", *Journal of the Royal Statistical Society* Series B, **24**, 406–424.

EDWARDS, A. W. F. (1972) *Likelihood* (Cambridge University Press).

MIZON, G. E. and HENDRY, D. F. (1980), "An empirical application and Monte Carlo analysis of tests of dynamic specification", *Review of Economic Studies*, (this issue).

PESARAN, M. H. (1974), "On the general problem of model selection", *Review of Economic Studies* **41**, 153–171.

PESARAN, M. H. and DEATON, A. S. (1978), "Testing non-nested non-linear regression models", *Econometrica*, **46**, 677–694.

QUANDT, R. E. (1974), "A comparison of methods for testing non-nested hypotheses", *Review of Economics and Statistics*, **56**, 92–99.

SARGAN, J. D. (1964), "Wages and prices in the United Kingdom", in Hart, P. E., Mills, G. and Whitaker, J. K. (eds.) *Econometric Analysis for National Economic Planning* (London: Butterworths).

SAWA, T. (1978), "Information criteria for discriminating among alternative regression models", *Econometrica*, **46**, 1273–1291.